The Stochastic Versus Deterministic Argument for Combat Simulations: Tales of When the Average Won't Do

Thomas W. Lucas

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943

Abstract

A continuing debate in the defense analysis community is whether combat simulations should (generally) be deterministic or stochastic. This paper argues that the nature of combat, along with fundamental mathematical principles, implies that most combat simulations should be stochastic. Although significant costs are associated with stochastic models, the resulting benefits will usually exceed the costs. For a given input set, stochastic simulations generate a sample of possible outcomes rather than a single result. Furthermore, deterministic approximations to stochastic elements often generate biases in outcomes, which might foster poor decisionmaking. To help focus an all-too-often abstract debate, this paper considers the spectrum of arguments for and against deterministic combat models. The emphasis is on real-world examples that illustrate that outputs of deterministic combat simulations tend to be biased. The examples cover the critical combat elements of attrition, detection, timelines, tracking, data fusion, and queues.

They aim at fixed values; but everything in war is uncertain, and calculations have to be made with variable quantities—Carl Von Clausewitz, On War

INTRODUCTION

Analysts use combat models to provide information to decisionmakers who must make and justify decisions involving billions of dollars and impacting many lives. For example, combat simulations, such as TACtical WARfare (TACWAR), have been used in many important Department of Defense analyses, such as the last Quadrennial Defense Review (QDR). Combat simulations tend to be extremely complex—large, nonlinear, and with numerous interactions. Models and their use in analyses can be simplified by making them deterministic. Examples of well-established and influential deterministic combat models include: the campaign simulation TACWAR, originally developed by the Institute for Defense Analyses (Joint Chiefs of Staff [1998]); RAND's Joint Integrated Campaign Model (JICM) & Simplified Tool for Analysis of Regional Threats (START) (Fox and Jones [1998], Wilson and Hagen [1996]); the Army's corps-level and campaign-level combat models Vector-in-Commander (VIC) and Concepts Evaluation Model (CEM); the Navy's campaign simulation Integrated Theater Engagement Model (ITEM); and the air-to-air combat simulation Advanced Air-to-Air System Performance Evaluation Model (AASPEM). Even more important, the Joint Warfare System (JWARS) reportedly will have a deterministic variant (JWARS Office [1998a]). JWARS is being developed as "a state-of-the-art…simulation of joint, campaign warfare" with application to "force assessment, planning and execution, system effectiveness and trade-off analysis, and concept and doctrine development and assessment" (JWARS Office [1998b]). It is important to understand the consequences of the deterministic approximations within these models. Moreover, it is necessary to consider whether we can improve our analyses—and, therefore, our decisions—by replacing the models' deterministic approximations with stochastic versions.

The primary objective of this paper is to alert model developers and consumers of simulation-based analyses to potential sources of bias associated with deterministic approximations to stochastic events in simulations containing multiple dissimilar entities interacting over an extended length of battle. The approach is to consider the pro and con arguments typically given in the deterministic versus stochastic debate. This provides structure to the debate. In addition, the list of arguments and the associated discussion may assist analysts and their customers in deciding which models to use, how existing models might be modified (or new ones built), and how to better interpret results.

DEFINING DETERMINISTIC AND STOCHASTIC MODELS

According to DMSO (1998): (1) A deterministic model is one "in which a given input will always produce the same output," and (2) A stochastic model is one "in which the results are determined by using one or more *random variables* to represent uncertainty about a process or in which a given input will produce an output according to some statistical distribution." The randomness can be induced by factors such as humans-in-the-loop; or, the random variables in stochastic models can be *pseudo random variables*—that is, the model is deterministic (and repeatable) given the random number seed values, Press et al. (1992). This paper focuses on the latter type of stochastic simulation.

There are two distinctly different classes of stochastic model. The first class involves sampling from a probability distribution of inputs that, once a sample of inputs is generated, the model is deterministic. In this situation, there are no random events within the simulated battle, i.e., there are no internal calls for a pseudo random number to determine the outcome of an event, such as a detection. With these models, a given probability distribution on inputs generates an empirical distribution on output measures. Davis et al. (1998) defines this approach as "probabilistic exploratory analysis." The second class contains those models in which events during the simulated course of battle are affected by the results of dynamically generated random numbers. Many of the points in this paper address both classes; however, the emphasis, particularly with the examples, is on the second class. Models, of course, can contain both a random sample of inputs and have internal random events. Finally, while some of the arguments may be applicable, this paper does not address optimization models, e.g., linear programs.

ARGUMENTS FOR MAKING COMBAT MODELS DETERMINISTIC

This section considers primary reasons, in the author's experience, that analysts give for developing and using deterministic models (or submodels). The reasons are roughly ordered according to their frequency of use. The discussion is necessarily general; in practice, the arguments must be adapted to the specifics of a given model and its use.

1. "A good point estimate is sufficient for my purposes"

This is the most common justification for using deterministic combat models. A good point estimate is frequently stated as the model providing (and the analyst using) "the most likely" or "average" outcome. The mathematical arguments and practical illustrations below show that a point estimate obtained from a deterministic combat model is probably biased. Thus, the model outputs are not likely to be good point estimates of either "the most likely" or "average" outcome.

It is often argued that one should not expect accurate point estimates from combat models. Instead, the models should be used to obtain insights and an understanding of general trends, e.g., Taylor (1983). The author agrees with this. Unfortunately, if a model's outcomes are sufficiently biased, the insights gleaned may be illusions that hinder rather than aid the decisionmaking process.

Another argument against strict reliance on point estimates is that good decisionmaking typically involves explicitly considering uncertainties—e.g., hedging strategies. For example, suppose that a particular course of action, called COA_i, has a 15-percent chance of a catastrophic outcome. A commander might reject COA_i even if the most likely and average outcomes for COA_i are highly favorable compared to the most likely and average outcomes of the alternatives. Models that generate distributions of outcomes facilitate explicitly addressing uncertainties in the decisionmaking process.

The New York Times (Stevens [1998]) recently recounted an example in which the reliance on a single value rather than on a distribution of possible futures resulted in costly and potentially catastrophic inaction. Stevens writes:

> *When the Red River … was rising to record levels in the spring of 1997, the citizens and officials of Grand Forks … had to rely on scientists' predictions about how high the water would rise. And in this case, … the flood forecast [by*

the National Weather Service that the river would crest at a record 49 feet] may have … made things worse…. Actually, there was a wider range of probabilities; the river ultimately crested at 54 feet, forcing 50,000 people to abandon their homes fast.… In fixating on the single number of 49 feet, the people … made a common error.… [The common error is rooted in a] tendency to overlook uncertainties, margins of error and ranges of probability [and] can lead to damaging misjudgments.

2. "Model elements are highly aggregated, thus the law of large numbers applies"

All combat models contain aggregations of some sort. Some believe that, when the aggregations are sufficiently large, a version of the law of large numbers (LLN) applies, thereby justifying the use of averages as good approximations to stochastic phenomena, e.g., Thomas (1997). Roughly speaking, the LLN states that for sufficiently large samples of independent and identically distributed random variables, the sample mean converges with near certainty to the true population mean (Feller [1971]). The LLN analogue for combat models is that if the units of aggregation (force sizes, time steps, etc.) are sufficiently large, then a stochastic outcome (e.g., attrition) is well approximated by its mean.

Taylor (1983), supported with additional references, found that the difference between the deterministic course of combat and the stochastic mean is small for (1) highly aggregated "simple" Lanchester models, with (2) sufficiently large forces, when (3) each side is willing to take substantial casualties, and (4) the sides are not near parity. The author's experience is that a LLN rarely, if ever, applies when the adversaries are competitive and the simulation is sufficiently complex to capture the dynamic nature of combat. A later section of this paper presents examples of highly aggregated models where the LLN does not apply.

Models that represent the dynamic and adaptive nature of combat are invariably nonlinear, including multiple discontinuities. It is shown below that problems arise when one propagates expected values in nonlinear models. Moreover, unpredictable regions of model space are typically those where one side or the other does not dominate—i.e., the forces are near parity (Hillestad, Owen, and Blumenthal [1995]). Of course, many of the important scenarios currently being studied involve situations where the sides are not near parity.

Another reason that a LLN usually does not apply is that there are often influential entities in simulations that exist in only a small number of discrete items. Influential simulation entities include aircraft carriers, special sensors (e.g., J-STARS), and command and control facilities. How does one account for the effect of .6 of an aircraft carrier? One can easily construct scenarios where the Blue force (a) wins easily if the carrier survives an attack, (b) loses if the carrier is sunk, and (c) has a mixed outcome if .6 of a carrier is modeled via its average sorties (i.e., 60% of a fully operational carrier's sorties), as done in ITEM (Science Applications International Corporation [1998]). The result of the battle may also depend on the times that the carrier could have been lost.

3. "Clever use of deterministic models can account for random elements"

In some applications sophisticated deterministic models account for stochastic variation. One approach to quantifying uncertainty with deterministic models is to analytically calculate output probability distributions. For instance, many formulas in probability modeling provide information on the probability distribution and/or moments of state space models (Ross [1985], Neimeier [1996]). Hence, it is unnecessary to estimate these distributions by stochastic modeling. Analytic derivations of probability distributions are particularly valuable when investigating low-probability events. The sample sizes necessary to precisely estimate low-probability events with Monte Carlo models are often prohibitive. Examples of deterministic approaches to calculate probability distributions include Hillestad (1982) and Yang and Gafarian (1995).

It is preferable, when possible, to generate an analytic (i.e., deterministic) solution to a model or submodel that provides information on the distribution of potential outcomes. Unfortunately, due to the complex and adaptive nature of combat, it is often impossible to obtain closed-form analytic solutions. This is especially true for large combat simulations containing stochastic attrition and adaptive decisionmaking. Horrigan (1967) shows that, even in relatively simple cases with adaptive combatants, "the simplest questions…when addressed in the combat environment, [can] entail equations of monumental complexity, well beyond the capabilities of probability theory and numerical analysis [to solve deterministically]."

Combat outcomes often depend critically on probabilistic "swing events." Some examples include: taking a strategic position (like Little Round Top at Gettysberg); first detection; environmental factors; and the loss of key elements, such as the fortuitous sinking of three of Japan's heavy carriers (Akagi, Kaga, and Soryu) within minutes during the Battle of Midway. Battle outcomes can qualitatively change depending on the random value taken by these vital factors of combat and usually cannot be sensibly averaged. Of course, analytic solutions can be conditioned on the outcomes of various swing events. Unfortunately, the number of potentially significant (different) trajectories through model space can increase exponentially (Hoeber [1981], chapter 1, and Horrigan [1967]). Furthermore, it is usually difficult to identify *a priori* all of the critical events upon which (and when in the battle) the results should be conditioned. In sum, while some applications can credibly be addressed through the clever use of deterministic models, this is very challenging with large simulations. One of the marks of a good analyst is to identify the key factors in the analysis (those that may affect the analytic conclusions) and focus the limited simulation runs to investigate them—whether deterministically or stochastically.

4. “Deterministic models require less processing”

Deterministic approximations to stochastic elements usually reduce processing requirements, and hence, model run time, by an order of magnitude or more. Monte Carlo simulations require more processing primarily because many replications are required to obtain precise estimates for a given input set, whereas a deterministic model needs only one run.

Additional processing is a cost one must incur when using stochastic models. It is one of the trade-offs that analysts must consider when building and/or using combat models. It is analogous to adding extra detail to a model: one should do it only if the analytic benefits outweigh the processing costs. The processing costs must also be put into perspective. In many analysis applications, the model runtime is a small fraction of the total research time. That is, the vast majority of time is spent on scenario generation, data acquisition, model modifications, testing, etc. Furthermore, processing replications in parallel and variance reduction techniques (Law and Kelton [1991]) can abate the additional processing costs.

5. “Deterministic models do not have to specify hard-to-quantify uncertainties”

> *Ignoring uncertainties is a chronic disease of planners, but to base a projection on a set of best guesses is absurd and can be disastrous—Quade and Carter (1989, page 162)*

Many of the numerous uncertainties associated with combat models are difficult to specify quantitatively. It is easier to specify a single value “best estimate” than a probability distribution. Unfortunately, a single value deterministic “best estimate” *is* a (degenerate point) probability distribution—one that is almost certainly wrong! In particular, a point distribution will understate the uncertainty. While assessing “the correct” distribution may be extremely difficult some simple forms are likely to capture much of the uncertainty. For example, if the distribution can be bounded above and below, and a most likely value selected, a triangle distribution might not be too bad of an approximation.

6. “Deterministic models are (usually) simpler”

> *Multiplicity ought not to be posited without necessity—William of Occam (Hughes [1997], page 22)*

Deterministic approximations allow us to replace random variables with single values. The extra parameters in stochastic models increase the analytic complexity by introducing additional factors that might affect outcomes and add random "noise" to model results. Of course, as an Associate Editor noted: "Sometimes it is harder to conceptualize and understand an averaging process rather than an explicit draw against a probability that more closely mimics the real world." In general, one should model to the level of detail necessary, and no more. Adding unnecessary parameters to a model costs in processing time and understanding. Analysts know this as the "curse of dimensionality" and it is very important (Hoeber [1981, chapter 1]). Good analysts keep models as simple as possible. If the additional parameters might significantly affect outcomes, then leaving them out may produce erroneous results. This is complicated by the fact that it is difficult to know *a priori* all of the factors that will be influential in combat models. One area that deserves special mention is decisionmaking by entities within the simulation. Stochastic models, since they may take many more diverse battle trajectories, often require richer, more complicated, decisionmaking algorithms.

7. "Stochastic models require statistical comparisons, which hinder understanding and complicate communication"

It is more difficult to find and explain patterns when random variation obscures the underlying signal, i.e., the limiting "model truth." Efron and Tibshirani (1993, page 1) write: "Left to our own devices we are not very good at picking out patterns from a sea of noisy data. To put it another way, we are all too good at picking out non-existent patterns that happen to suit our purposes." When the signal we are looking for is hidden by random noise, we must use statistical methods to determine whether or not the patterns we see are readily explainable by random variation. That is, are the apparent patterns probably really true?

Statistical comparisons (e.g., the probability that system A is more survivable than system B is at least .80) are not as satisfying as deterministic comparisons (e.g., system A is more survivable than system B). Furthermore, with stochastic models there is a positive probability of making erroneous conclusions purely by chance, i.e., statistical Type I and Type II errors. Unfortunately, the nature of combat, with all of its uncertainties, is such that probabilistic comparisons are sometimes the best we can credibly provide. It may be appropriate to suppress (some of) the uncertainties when communicating the results; however, one should always do the analysis correctly (including stochastic components, if necessary) and then communicate the results appropriately. Excellent guides on effectively communicating (statistical) data include Cleveland (1994) and Tufte (1983).

8. "A deterministic model is sufficient to support how I am using the model"

All models are wrong but some are useful—George Box (1979)

It has been argued that all aggregate-level combat models (deterministic or stochastic) are not valid in the sense that their outcomes can be considered accurate predictions of potential future real-world events (Hodges and Dewar [1992], Oreskes, Shrader-Frechette, and Belitz [1994]). Nonetheless, unvalidated combat models can credibly be used in an analysis to generate hypotheses, learn about the consequences of assumptions, support *a fortiori* arguments, aid in communication, etc. (Hodges and Dewar [1992]). That is, a good analyst can provide useful information, even when the model has substantial biases. There are plenty of clear, concise, and insightful analyses based on simple deterministic models (e.g., Hughes [1995] and Dupuy [1990]). Therefore, deterministic combat models can play a credible role in analyses. However, that role can be complicated by an accumulation of biases due to the deterministic approximations. This is particularly true for larger, more complicated models than those just cited, such as TACWAR. For most purposes, we are better off if we can mitigate the biases.

ARGUMENTS FOR MAKING COMBAT MODELS STOCHASTIC

This section considers primary arguments in favor of stochastic combat models (or submodels). These generally apply to all of the elements within a combat model that could (or should) be modeled stochastically. While all of the reasons are important, the first two are especially decisive.

1. Many events in combat are intrinsically stochastic

Combat is inherently stochastic! Ancker (1995) states that "stochastic target attrition processes" are *axiomatic* to combat. Of course, when one side dominates an engagement the stochastic variation may be insignificant, e.g., Desert Storm.

In combat, a vast number of factors with complex interactions affect battle outcomes. Many of the factors are impossible to precisely determine, i.e., reasonably modeled with single values rather than probability distributions. Because this truth is central to this paper, we expand on it in this subsection. There are several distinct causes of uncertainty in combat models, see

Hodges and Pyles (1990) for a broader discussion. We will look at four classes of uncertainty (not necessarily mutually exclusive) in combat analysis as defined by Youngren (1998).

Outcome Uncertainty

Youngren defines *outcome uncertainty* as "uncertainty related to the random nature of the underlying processes…[with current knowledge and tools]." Many elements in combat fit this category, including fundamental ones such as attrition, detection, and the environment.

Modeling attrition has been and continues to be a vexing problem for defense analysts. Even one-on-one, high-resolution, physics-based models have difficulty predicting single-shot engagement outcomes with known precision. Consider the detail needed to accurately model the effects of an air-to-air missile homing in and detonating near a maneuvering aircraft. Perhaps a single fragment perforates a fuel tank, causing a slow leak that results in a catastrophic fire minutes after the blast. James F. O'Bryon (1996), the director of the Department of Operational Test and Evaluation's Live Fire Testing writes: "Literally, not one purely physics-based model exists in the nation to predict conventional weapons effects on a platform." How, then, can we expect deterministic multi-platform models to produce accurate point estimates?

It is also impossible to model detection events precisely. The Defense Science Board (1989) reports that when "[six] experts…[with] long-term experience with over-the-horizon radars," were asked to independently calculate from their models—under identical circumstances—the target size for a 50-percent probability of detection at 1200 miles, their answers varied by more than a factor of ten. The sensitivity of sensors to unpredictable environmental effects further complicates matters.

Attrition and detection events can be estimated statistically using field data, e.g., operational, test, or training exercises. Of course, field data estimates all have their own uncertainties—often quantified with confidence intervals.

Process [and Object] Uncertainty

Following Youngren, *process (and object) uncertainty* is "related to our inability to understand an object or process completely. [It] arises when we try to model something for which information is not available, or the process itself is not well understood." The prototypical example of this in combat models is our lack of complete knowledge about threat systems and tactics. Inevitably, we must rely on uncertain intelligence assessments. Unfortunately, experienced analysts know our models' outputs can be extremely sensitive to how enemy performance and tactics are modeled. Additional examples are "provided by all other processes which we do not understand; e.g., suppression, morale, and leadership."

Future Uncertainty

Youngren defines *future uncertainty* as "uncertainty related to the future setting of events or processes." Of course, making accurate predictions is hard, especially about the future. While it has always been challenging to predict the future in defense analysis, the fall of the Soviet Union has made this truer than ever. Critical factors that are more uncertain than ever include the opposing force(s), future battlefields, potential coalition partners (who, in 1989, thought the United States could be coalition partners with Syria?), warning time, and threat (asymmetric) strategies (Garner [1997]). There is also tremendous uncertainty about the future systems that will be acquired and their operational performance—for both our own systems and those of our potential coalition partners and adversaries.

Decision Uncertainty

Youngren defines his last uncertainty, *decision uncertainty*, as "uncertainty related to decisions." Even when decisions are made by strictly objective criteria, e.g., by computer algorithms, Williams (1986) shows us that optimal decisionmaking sometimes requires that "the decision regarding strategy [depend] entirely on some suitable chance event." For this paper, we expand decision uncertainty to include uncertainty in other human elements, such as endurance and will to fight. Simply put, there is tremendous variability in decisionmaking among both individuals and groups, and these differences can be decisive in combat. Failure to account for the ability of leaders to make wise (or poor) decisions and to inspire their troops may result in very poor estimates of the effects of information and other elements on battle outcomes.

The effects of leadership and troop morale are very difficult to model. Dupuy (1985) notes that it has been said (perhaps by Wellington or Blucher) that "the mere presence of Napoleon on a battlefield was worth 40,000 men." Furthermore, the morale of troops, their effectiveness, and their willingness to continue to fight effectively under extreme adversity are highly variable and difficult to predict (again, see Dupuy [1985]). Bankes (1996) quotes General Ronald Fogleman:

> *I challenged the Air Staff to model the Desert Storm air campaign after the fact when we had all the information available. So, with perfect 20-20 hindsight, I*

asked them to find out why the huge casualty predictions did not *come true. They said they couldn't do it because they determined that the Iraqis acted so irrationally that their actions could* not *be modeled.*

It is interesting to note that while most of the pre-Desert Storm analyses significantly overestimated the time required to win the battle and the casualties of the United States and its allies, Dupuy (1990, page 131), using a very simple deterministic model, was qualitatively better than many of the more sophisticated analyses. Furthermore, analysts could have, even should have, addressed the uncertainty in Iraqi resolve with excursions rather than by sampling from a distribution. This type of exploration is particularly useful in addressing future uncertainty. One nice feature of having to put a distribution on a parameter, as in stochastic models, is that it automatically generates those excursions.

There is tremendous uncertainty associated with combat modeling, and our answers can be sensitive to how we treat the uncertainty. To ensure that we give decisionmakers the best information possible, we must address uncertainty in our analyses. Burying our vast uncertainty with single-value estimates is a recipe for self-deception. Traditional sensitivity analysis (SA) can help somewhat; however, combinatorics usually prohibit comprehensive SA (Dewar et al. [1996a]) and for many analysis purposes it is simply insufficient to determine how locally sensitive a biased answer is. Quade and Carter (1989, page 163) write: "In a better approach to sensitivity testing we use a Monte Carlo sampling process to examine sensitivity to the simultaneous variation in a number of parameters."

To address the uncertainty inherent in combat, we need methods that help us quantify uncertainty. One can implement (in most cases) two approaches to quantifying uncertainty with Monte Carlo techniques: (1) by obtaining statistics from experimental data, and (2) by assessing subjective expert-based probability distributions on unknown quantities. While the former is preferred, for many of the uncertainties discussed above the latter is often the best we can do. Approaches for assessing subjective probabilities, and subsequently updating them upon observing data, can be found in texts on Bayesian statistics, e.g., Berger (1985) and Berry (1996).

Bayesian statistics provide a consistent theory for quantifying all uncertainty through probability assessments. Prior to performing the modeling experiment, our uncertainty is quantified into prior beliefs, which are updated by experimental data. Bayesian methods are particularly useful when there is a paucity of data and a plethora of expert knowledge—which characterize many defense analyses. While most of Bayesian theory is personalistic, there are several methods for building "objective" priors, sometimes referred to as reference or public policy priors (Bernardo and Smith [1994], Press [1989]). Assessing subjective probability distributions is difficult. In situations where it is difficult to quantify uncertainty, several priors can be studied. In addition, there are cases where the uncertainty cannot be reliably specified, but can be bounded, e.g., the uncertainty is uniformly less than that of a given distribution. Another approach is to use as a prior the distribution that minimizes our information (or maximizes the entropy) constrained by what is known (Jaynes [1983]), e.g., boundaries and moments. Sophisticated analyses can put a probability distribution over the potential priors; this is called hierarchical modeling (Good [1980]).

2. Propagating expected values causes biases in nonlinear functions

Combat models, reflecting the nature of combat, are usually highly nonlinear with numerous discontinuities, such as thresholds. Unfortunately, if F is a nonlinear function (e.g., a combat model), X a random variable, and E the expectation operator, then, $E[F(X)] \neq F(E[X])$. The consequence of using E[X] as a substitute for the random variable X is a bias in the estimate of E[F(X)] (Savage [1998, Chapter 3]). Most of the examples in the next section are variations on this theme.

A familiar version of the fact that expected values do not propagate well in nonlinear functions is Jensen's Inequality, which states that if F is convex, then, $E[F(X)] \geq F(E[X])$. Note: A function F[.] is convex on an interval (a, b) if for all λ [0,1] and each x, y (a, b) $F[\lambda x + (1-\lambda)y]$ $\lambda F[x] + (1-\lambda)F[y]$ (Royden [1968], page 108). More generally, by a Taylor's series expansion of a model F on a random variable X (equation [1]) we see that "variance matters" in the expectation of F(X).

[1] $$\mathbf{E[F(X)]} = \mathbf{F}(m_{\mathbf{X}}) + \frac{s^{\mathbf{2}}}{\mathbf{2}}\mathbf{F}^{(\mathbf{2})}(m_{\mathbf{X}}) + \mathbf{E[R_n]}$$

where: $\mu_X = E(X)$

R_n = a remainder term.

It follows that, provided the model is nonlinear, using "best estimates" in place of distributions will result in combat models generating biased outcomes. Of course, the size of the bias depends on the degree of nonlinearity. Koopman (1956) identifies *linearitis* (the assumption that every function is linear) as one of the major fallacies in operations research.

3. It is often better to reason probabilistically

A previous section discussing why point estimates, even if accurate, are insufficient for many purposes addressed this issue. Understanding a distribution of potential outcomes, particularly the tails of the distribution, is especially important in the current risk-averse environment. Decisionmakers may be more concerned about the risk of significantly undesirable outcomes than mean outcomes. Many of the recent missions, e.g., Somalia, are situations in which the probability of sustaining significant losses is an important factor in the decisionmaking process.

4. Stochastic elements tend to reduce instability in combat models

Analysts have long been aware of stability problems with combat models. Some good references on this are Dewar et al. (1996b), Sandmeyer (1990), and Saeger (2000). One peculiar phenomenon is called non-monotonicity, which occurs when increasing one variable's value (such as the number of Red combatants), while keeping all other variables constant, results in poorer outcomes for Red even though one expects better outcomes. Louer (1993), Huber and Tolk (1994), and Lucas (1997) have argued and demonstrated that adding stochastic elements to deterministic combat models tends to smooth model outputs. Intuitively, replacing single points with probability distributions averages out some of the hard "break points" in models. However, this line of research shows that randomization does not always eliminate or even reduce non-monotonicity (Allen, Gillogly, and Dewar [1993]), and the choice of where to put the randomness and how much to add can have substantial consequences (Lucas [1997]).

5. Monte Carlo models explore more regions of feasible model space

For a given set of fixed inputs, Monte Carlo models, by necessity, are sampled many times. Each of the samples takes a different trajectory in the simulated battle. Thus, for a given input, multiple Monte Carlo runs explore more of the feasible model space than a single "mean course of battle" deterministic run.

The surprise insights we obtain when using models are often the most informative. Expanded model exploration increases the opportunity to acquire surprise insights. Some surprises occur only in extreme regions of model space and/or when certain interactions occur. For example, a badly overmatched Red force might win a battle when it detects Blue at a distance of two standard deviations above its expected detection range, and Blue's defensive systems operate at one standard deviation below their expected effectiveness. This effect will not be discovered with a deterministic model that uses only the means of these random variables. In addition, an excellent approach to finding errors in models is to extensively exercise them. Monte Carlo methods typically generate many more diverse trajectories through a simulation model's space, thus increasing the chance of catching software and logic errors.

6. When you fix stochastic elements you typically under-estimate the total variability

Most combat models contain numerous random elements that contribute to the overall variance in outcomes. While not universally true, the total outcome variability generally increases with the number of stochastic elements modeled. A simple, yet illustrative, example of this is that the variance of a sum of independent random variables is the sum of the variances. Similarly, when some stochastic elements are treated deterministically, estimates of output variability are usually low. In the extreme, a completely deterministic model provides no empirical estimate of variance. Inferences about tail behavior, such as estimates of maxima and minima, are error-prone when uncertainties are understated.

7. Monte Carlo models facilitate dealing with uncertainty and complexity

The complexity of all but the simplest combat models makes analytical solutions prohibitive, particularly if uncertainties are represented by probability distributions. Fortunately, the structure of many models allows us to obtain a random sample of potential outcomes by Monte Carlo methods. As Hammersley and Handscomb (1964) write: "The idea behind the Monte Carlo approach…is to [replace] theory by experiment whenever the former falters."

EXAMPLES OF BIASES RESULTING FROM DETERMINISTIC APPROXIMATIONS OF STOCHASTIC EVENTS

The above arguments provide a list of things to worry about when using deterministic approximations to stochastic factors in combat models. It remains to be shown that these are of more than academic concern. Towards that end, this section will examine a series of examples in which deterministic approximations yield biased results. These examples cover many important elements in combat modeling, such as timelines, attrition, detection, tracking, data fusion, and queues. All the examples are stylized versions of issues that have arisen in combat analyses. While sometimes simplified for expository purposes, they are fundamentally equivalent to actual examples that could have (or did) result in models generating misleading information. Most of the simulations in use today contain calculations similar to some of the examples.

Timelines

Many events in combat simulations are characterized by the time it takes to complete a task or series of tasks. Examples of time delays include: the time that it takes a sensor to slew and acquire a target; the time a person (or computer) uses to assimilate data and update his (or its) perception and/or make a decision; the time it takes a message to propagate through a congested network; and the time required for equipment to be deployed into a theater. In most cases, far too much detail is required to accurately model—if it is even possible—these events via physics-based methods. For many purposes, the single value (or probability distribution) of event time sufficiently characterizes the event or process being modeled. As a consequence, these events are often modeled by fixed time delays. Inevitably, there are uncertainties that cannot be rectified, e.g., natural variation and unforeseen exogenous factors.

A series of time delays often precedes a decisive event, such as a weapon firing. For example, consider the timeline in Table 1 for a point-defense anti-air missile system defending against an incoming cruise missile. The incoming missile must be detected and subsequently acquired (track established), identified, and designated to an illuminator and launcher; the illuminator must slew and lock-on the target; and the defensive weapon(s) must be launched (we assume a salvo size of 2). All of these events (except salvo delay time) are subject to random delays. For simplicity, we assume in Table 1 that all of the times follow Gaussian distributions (denoted $N(\mu,\sigma^2)$), are independent, and we ignore truncating for infeasible negative times. That is, we simply sum the variances.

Table 1: Deterministic and probabilistic timelines result in qualitatively different estimates of point defense system effectiveness.

Event	Approximate distribution	Time remaining for intercept	Expected value propagation
Detect:		N(17,5)	17
Acquire:	N(3,1)	N(14,6)	14
Identify:	N(4,4)	N(10,10)	10
Designate:	N(2,1)	N(8,11)	8
Illuminator Lock:	N(4,4)	N(4,15)	4
Launch Delay:	N(2,1)	N(2,16)	2

Salvo Launch Delay:	1.5	N(.5,16)	.5
PK with .8 missile		**.64**	**.96**

When we propagate expected values we always get two good shots at the incoming missile, with a probability of kill of $1-(1-.8)^2 = .96$. In the stochastic version there is a .31 probability that no shots are made, a .14 probability that only one shot is taken, and a .55 chance that the ship can get off two shots; with a resultant probability of kill of .64. We get qualitatively different answers in Table 1 depending on whether we propagate the probability distribution of the accumulated timeline or just the mean. Propagating mean delays results in underestimating the probability that the threat leaks by a factor of nine (.04 versus .36). Here, the problem is structured so that it is analytically tractable. In many cases, the distributions are not so friendly (e.g., significant truncation effects exist; the delay times are not independent; and/or the distributional form of the sum is intractable). Fortunately, even with complex distributions, we can often simulate processes like this with Monte Carlo methods.

While this illustration is simplified for clarity, the essential features are applicable to more complicated models. Furthermore, delays such as this are ubiquitous in combat models. When simulating time delays, it is usually better to use a probability distribution than a fixed number. The distributions can be estimated by empirical data, theoretical methods, or more detailed submodels. In this example, each time we propagate a deterministic value, we understate the variance of the accumulated delay, and the bias increases. Typically, modeling some, but not all, of the elements stochastically reduces, but does not eliminate, the biases associated with deterministic models. In addition, in this example, using the "average" detection time also affects the bias.

Attrition and decision thresholds

Attriting enemy forces, or positioning oneself to do so, is fundamental to combat. Combining Ancker's (1995) two axioms of combat yields: "all combat is a hierarchical network of firefights…with stochastic attrition." As discussed above, we cannot even faithfully model the effects of conventional munitions against a single platform (O'Bryon [1996]). The situation is even more complicated for scenarios involving many-on-many. Another complicating factor is dynamic decisionmaking. That is, based on the evolving state of battle, decision entities decide when to call in reinforcements, where to move, how to attack, whether to proceed or withdraw, etc. These decisions often depend, in part, on the outcome of attrition as it affects force levels and force ratios. Consequently, they are treated together here.

Battles evolve in dramatically different ways depending on (random) attrition, as the following example illustrates. Ten helicopters go on a hazardous deep rescue mission with a staging stop. The assumed reliability of each helicopter is .7 and is independent of the others. Six functioning helicopters are needed at the staging stop to carry out the mission. Using expected values, seven helicopters are available at the staging stop, and the mission always continues (with one helicopter to spare). In the stochastic version, there is a 15-percent chance that five or fewer helicopters are available at the staging stop, and the mission is canceled. If the cost of aborting the mission is sufficiently high, a 15-percent chance of failure can be decisive in deciding whether or not to attempt the mission—or in determining how many helicopters are required. Failure to account for the probability of having to cancel or otherwise modify missions due to attrition can result in models producing misleading information—and in bad plans being implemented. Many elements in combat models are similar to this example, such as bombers penetrating SAM belts, submarines passing through anti-submarine screens, and tanks and soldiers advancing through minefields or enemy fire.

We now turn to a more complicated model that has been extensively studied in recent years, the so-called "Dewar model" (see Table 2). Dewar, Gillogly, and Juncosa (1996b) used this model to study stability issues in combat models. The baseline deterministic Dewar model contains 18 parameters, a very "simple" and highly aggregated model by modern combat modeling standards. Parameters in the model include initial force sizes (B_0, R_0), reinforcement delays, reinforcement levels, total reserves, attrition rates, and decision thresholds. At each time step, each side's attrition is a linear function of the other side's engaged forces (B_n, R_n)—i.e., the model is a difference equation variant of Lanchester's (1916) "aimed fire" model with fixed coefficients. Furthermore, at each time step, depending on the engaged force ratio and force levels, each side makes decisions about withdrawing or calling in reinforcements. If summoned and available, reinforcements arrive in fixed numbers after a given delay. The essential outcome measure is a binary decision of who wins, Red or Blue.

Table 2: Dewar's simple combat model (Dewar, Gillogly, and Juncosa [1996b]).

	Blue	Red
Initial troop strength	$B_0 = 500$	R_0

Combat attrition calculation	$B_{n+1} = B_n – R_n/2048$	$R_{n+1} = R_n – B_n/512$
Reinforcement thresholds	$R_n/B_n \geq 4$ or $B_n < 0.8\times B_0$	$R_n/B_n \leq 2.5$ or $R_n < 0.8\times R_0$
Reinforcement block size	300	300
Allowable reinforcement blocks	5	5
Reinforcement delay(time steps)	70	70
Withdrawal thresholds	$R_n/B_n \geq 10$ or $B_n < 0.7\times B_0$	$R_n/B_n \leq 1.5$ or $R_n < 0.7\times R_0$

One of the main conclusions from Dewar et al.'s (1996b) research is that, even in this simple model, there are regions of model space that are highly non-monotonic, even chaotic. Figure 1 displays a one-dimensional slice through the Dewar model. Battle outcomes are non-monotonic (i.e., there are many reversals of outcome) as the number of Red forces initially engaged (R_0) ranges from 890 to 1600, with all other variables held constant. In this region, the model suggests, implausibly, that adding more Red troops (almost double) can be bad for Red.

Non-monotonicities in combat models are not a rare phenomenon (Cooper [1994]). Non-monotonicity may be a real underlying phenomenon, as Davis (1992) discussed, or the result of a poorly formulated model—in particular, with respect to decisionmaking, (Cooper [1994], Huber and Tolk [1994]). Louer (1993) and Lucas (1997) have shown that making deterministic simulations stochastic can often dramatically mitigate non-monotonicities. We will focus on this latter issue. Let's compare what happens to the Dewar model when all eight of its thresholds (two withdrawal and two reinforcement thresholds for both Red and Blue) and attrition are stochastic.

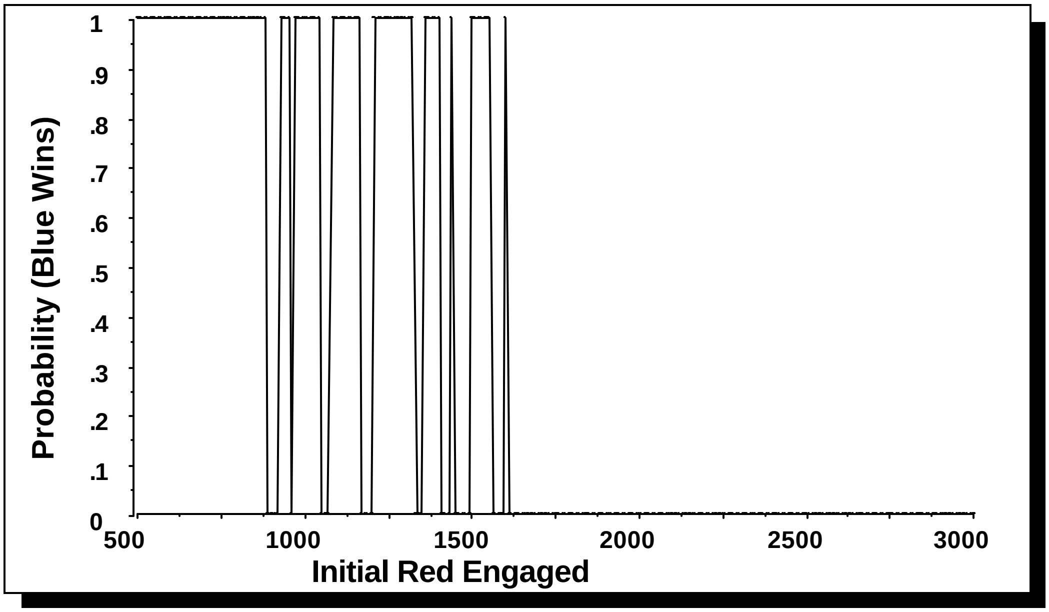

Figure 1: There are many reversals in outcome (i.e., non-monotonicity) in the Dewar model as the number of Red forces initially engaged increases (Lucas [1997]).

Figure 2 shows the estimated probabilities that Blue wins when:

(1) All eight decision thresholds are randomly chosen prior to each run from a uniform random variable ranging from 20 percent below to 20 percent above the threshold's value in the baseline Dewar model.

(2) At each time step within a run, the attrition is sampled from a truncated normal random variable with the means taken as the deterministic values in the Dewar model and standard deviations equal to the square root of the means—as occurs when attrition follows a Poisson distribution.

This requires only two additional parameters. One thousand replications are taken at each force level to ensure that the uncertainties around the estimates are negligible. Three things are immediately clear: (1) there is no significant non-monotonicity; (2) the trend appears reasonable; and, (3) the deterministic mean course of combat is dramatically different from the stochastic one. The stochastic variation averages across a range of battle trajectories; thus, neighboring input points do not follow dramatically different branches with certainty. As a consequence, in the stochastic case of Figure 2, even if the numbers are wrong, we get a consistent ordering of the alternatives (i.e., more Red troops doesn't hurt Red); whereas we didn't in the deterministic runs in Figure 1.

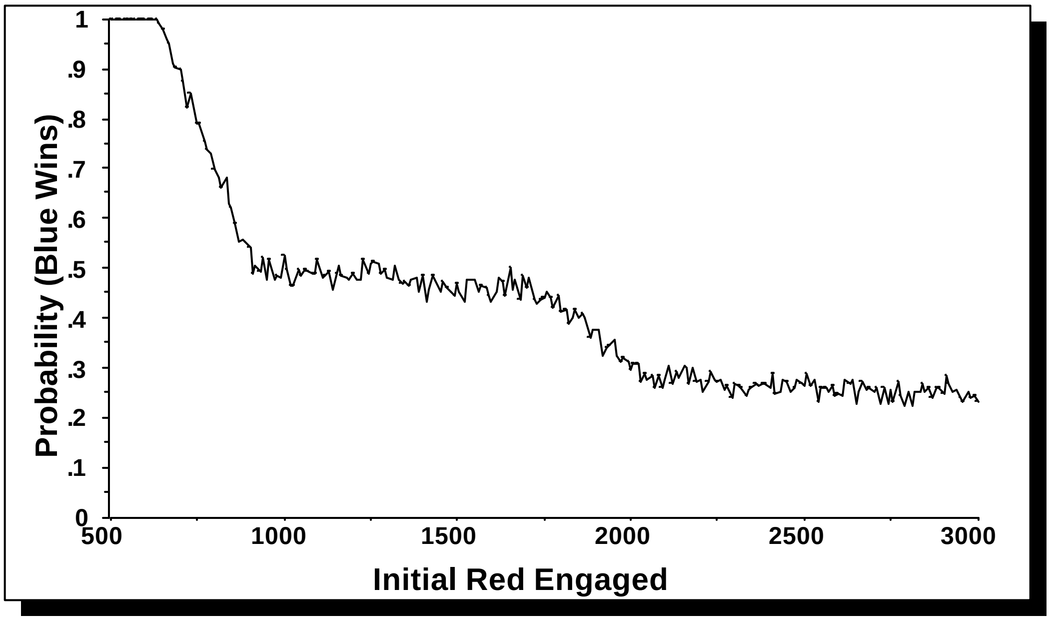

Figure 2: Same set-up as Figure 1, except now the (eight) thresholds and attrition are stochastic (Lucas [1997]).

We have reviewed two examples where stochastic attrition (and stochastic thresholds based on force levels and ratios) yields qualitatively different results from their deterministic counterparts. Additional articles that show how stochastic simulations differ from deterministic ones include: McCue (1999), Youngren (1998), Hillestad, Owen, and Blumenthal (1995), Hofmann, Schnurer, and Tolk (1995), Johnson, Isensee, and Allison (1995), Concepts Analysis Agency (1991), Ancker and Gafarian (1987), and Taylor (1983). These papers report varying degrees of divergence between the deterministic mean course of battle and the equivalent mean of the stochastic versions—in most cases with highly aggregated models. The divergence is generally largest when the forces are near parity.

Detection

[Detection] is in general a random event—Bernard Koopman (1980, page 51)

Obtaining early and accurate information about opposing forces can be decisive in combat. In combat models, information about the other side is usually obtained by model objects that simulate the detection capabilities of various sensors, such as radar, sonar, infrared, and visual systems. The algorithms that model sensors span the resolution spectrum from cookie cutters to detailed pulse-by-pulse ray tracing. Unfortunately, as was found by the Defense Science Board (1989) when comparing the results of six simulations, under identical initial conditions, even the highest resolution models can disagree dramatically. Consequently, modeling sensor detection performance deterministically runs the risk of biasing outcomes. As we saw in the timeline example above, a failure to include the stochastic variation in detection range may result in biases in estimates of system effectiveness—in that case, ship self-defense.

In the following example, gleaned from Bennett et al. (1998), fixed detection ranges produce erroneous outcomes in a more complex situation—arising in a mission-level aircraft survivability study. While examining the effect of stealth on aircraft survivability the authors noticed that their model suggests that sometimes more stealth, with everything else held equal, reduces aircraft survivability (see Figure 3).

The authors determined the subtle cause of this implausible result to be a function of the complex surface-to-air missile (SAM) "engagement footprints" and deterministic detection and engagement processes. The explanation is best understood by examining Figure 4. The analysis used two models: (1) a high-resolution one-on-one engagement model that generates detailed "engagement footprints" between SAM sites and aircraft signatures, and (2) a mission-level simulation that models multiple SAM types at various locations, the command and control system that links them, and multiple aircraft. Using engagement footprints generated in the engagement-level model, the mission-level model calculates the survivability curve in Figure 3.

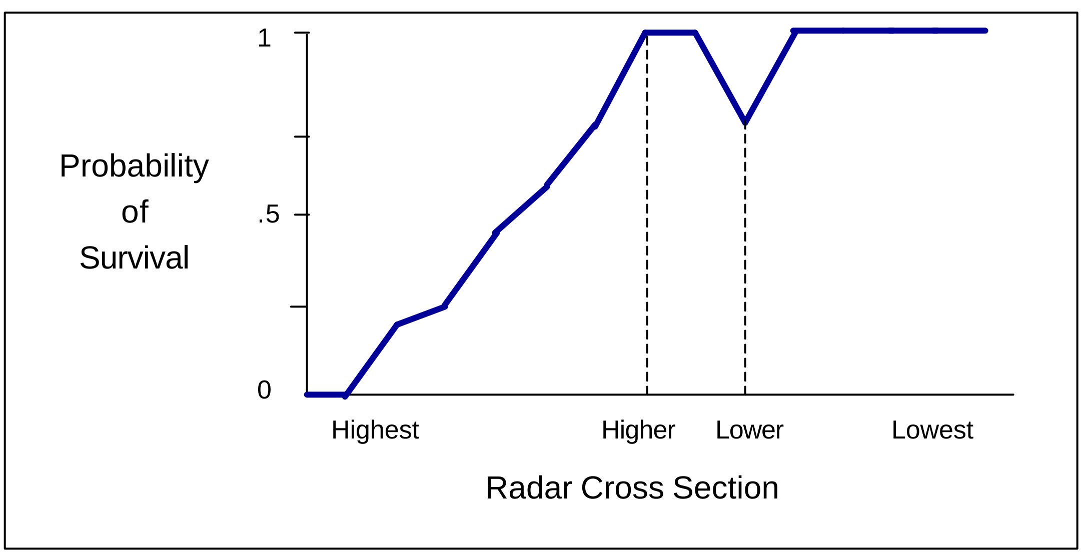

Figure 3: Using a deterministic detection range an aircraft with the "Higher" radar cross section is more survivable than one with the "Lower" radar cross section (Bennett, Hagen, Chow, Brooks, Myrick, Hobbs, and Gates [1998]).

The footprints in Figure 4 capture the defended area of a surface-to-air missile. They may be interpreted as a SAM site in the center of the footprint (at the cross) with an aircraft flying from the top of the figure to the bottom of the page at various offset angles. The darker areas correspond to intercept points where the SAM will intercept the aircraft with a high probability of kill and the light areas are where no intercept occurs.

In this example, the engagement footprint for the lower RCS aircraft is smaller than the engagement footprint for the higher RCS aircraft. Why, then, do we get this nonintuitive result? Bennett et al. (1998) write:

> *In the first case, the higher RCS aircraft is tracked with delays that cause the SAM to take the first shot in a small area of high survivability. Not only is this aircraft fortunate to avoid being killed by the first shot, a second shot also occurs in an even smaller area of high survivability. The lower RCS aircraft is not as fortunate. Because of its lower cross section, it is tracked a little later than the higher RCS aircraft (as it should be). Unfortunately, this delay draws the lower RCS aircraft into a more deadly part of the engagement footprint. The first shot results in a kill.*

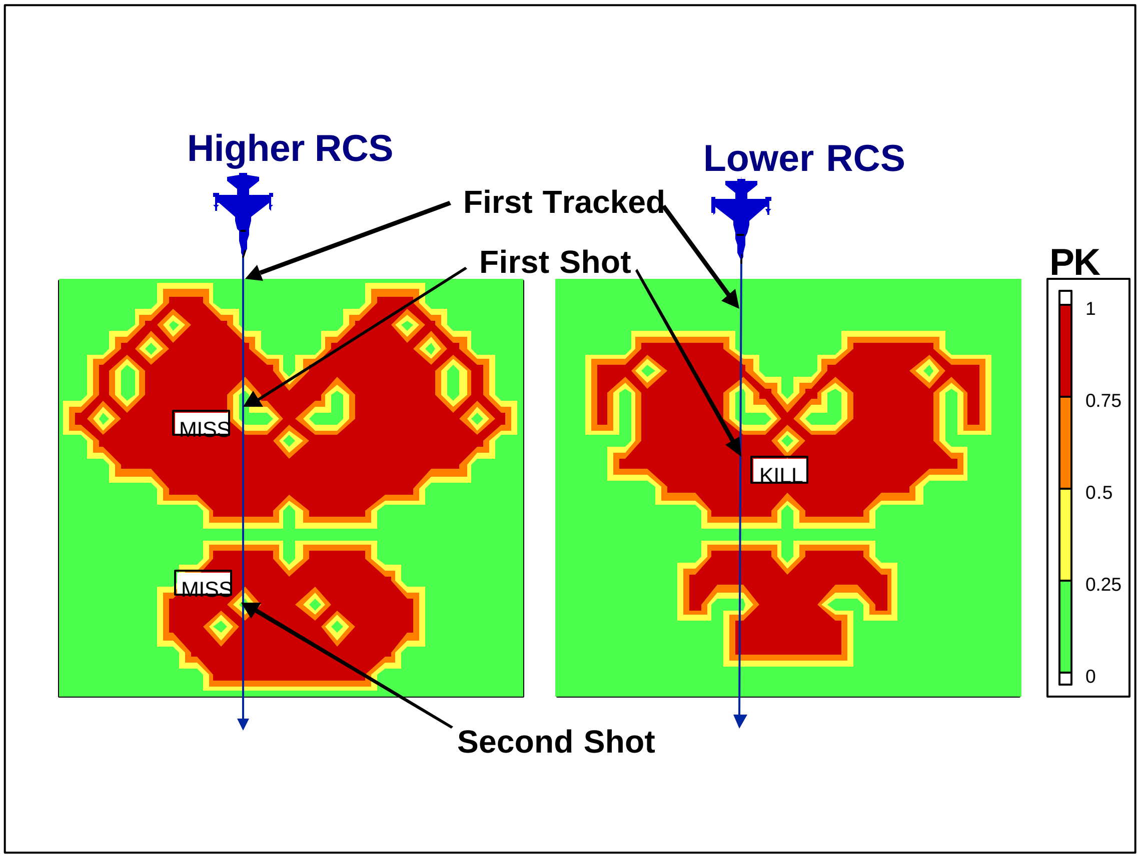

Figure 4: For some geometries, with a fixed detection range and engagement process, an aircraft with a lower radar cross section (RCS) is less survivable. In this example, the aircraft with the higher RCS (left panel) is fortunate to have the intercepts always occur in areas of low PK. The lower RCS aircraft, while detected later, is not as fortunate (Bennett, Hagen, Chow, Brooks, Myrick, Hobbs, and Gates [1998]).

Their deterministic modeling of the detection, tracking, and engagement processes results in an aircraft, of a given signature and trajectory, always having the same engagement sequence—that is, always falling into the SAM miss "holes." In the field, of course, detection ranges and engagement processes vary considerably, depending on the environment, system, and operator differences. Therefore, there is a distribution of engagement sequences that causes an averaging of the effects and a smoothing over of the footprint "holes." When the authors modeled these events stochastically, via Monte Carlo methods, they found that "the lower RCS aircraft survives more, rectifying the non-monotonicity observed [in Figure 3]."

Sensor target location and system tracking errors

Once threats are detected, target tracks are created and updated as new information is acquired. Track files contain estimates of multiple track attributes, typically including identity, position, heading, and velocity. This information is vital in assessing threats and choosing options to counter them. Due to uncertainties, such as sensor measurement errors, the track estimates are imprecise. This results in errors when computing threat intentions (e.g., the intended target of an incoming missile) and in the effectiveness of potential counter-measures (e.g., the probability that a defensive weapon can acquire and kill an incoming threat). Consequently, measures are taken to increase track accuracy, such as statistical algorithms that combine and smooth the detection information (Blackman [1986]).

Deterministic combat models often allow model entities access to a target's "ground truth" information—which typically corresponds to the mode of the probability distribution of the target track. Failure to account for tracking uncertainties (i.e., allowing some model entities to unrealistically possess the actual identity, position, heading, etc. of other model entities) leads to erroneous estimates of system effectiveness.

Using the Evolutionary SAM SIMulation (ESIM), Lucas, Schwab, and Feldman (1990) studied the sensitivity of a distributed medium-range surface-to-air missile system's (MSAM) effectiveness to sensor error and tracking performance. ESIM was enhanced by adding tracking errors through lookup tables. The study scenario involved a networked Blue force consisting of five ground-based sensors, nine combat operational centers, and 36 SAM launchers with eight ready missiles each. The attacking Red force was composed of 38 fighter-bombers with four ground-suppression missiles each.

Figure 5 shows how several mission effectiveness measures vary as a function of steady-state track error. The areas represent the means of 30 Monte Carlo trials at multiple system track accuracies, ranging from perfect tracking to very poor tracking. The three areas correspond to the average number of Blue SAMs that kill, acquire but miss, and fail to acquire an attacking bomber. The sum represents the total number of SAMs fired. The estimated standard errors of the sample means are around one.

As the tracking error increases targeted threats are less likely to be "in the basket" of defensive SAMs. Therefore, the ability of the intercepting missile to successfully acquire the targets varies considerably as a function of track accuracy. In this example, the Blue force has sufficient firepower so that the effect of poor tracking is primarily an increase in weapons expended. In more stressful scenarios, however, the important sensitivity is with respect to threat survival—and correspondingly the threat's ability to inflict losses on Blue. Failure to account for sensor/tracking errors yields biased estimates of the system's effectiveness. Monte Carlo techniques, while not the only approach, facilitate the modeling of sensor and track errors.

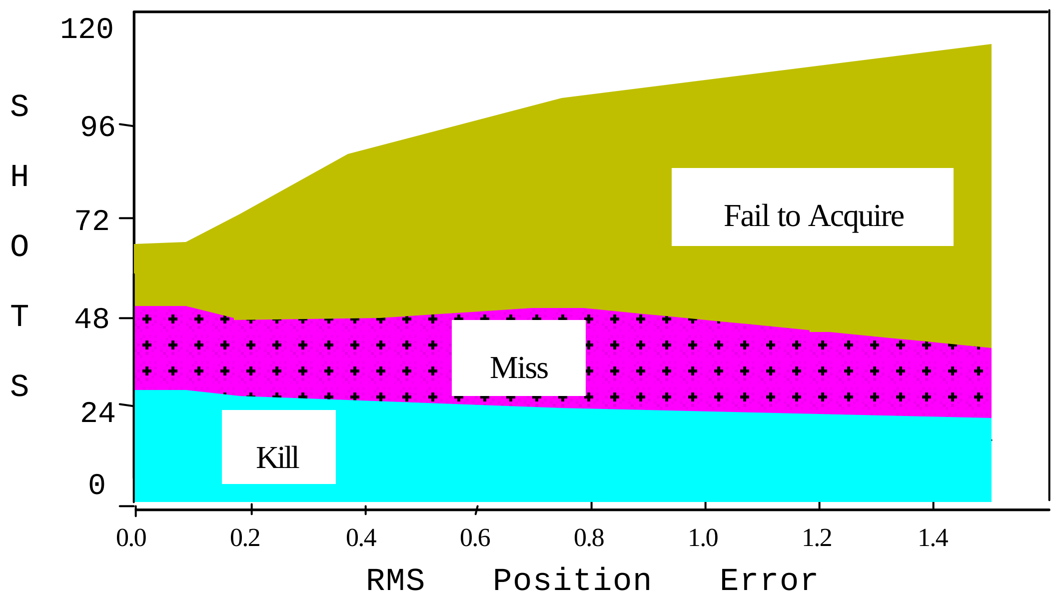

Figure 5: The sensitivity of mission effectiveness to track accuracy in a distributed land-based area anti-air warfare (AAW) scenario (Lucas, Schwab, and Feldman [1990]).

Data fusion

On modern battlefields, command and control centers combine information from multiple sources, referred to as *data fusion*. This is done to obtain the most complete and accurate understanding of the situation possible—i.e., to maximize battlespace awareness. Specifically, observations and tracks from various sensors and command posts are combined. One operational challenge is to determine whether multiple tracks or observations constitute multiple reports on the same entity or distinct information from more than one entity. These decisions are called *track correlation* or *track association*. These assessments affect command and control decisions, and are, of course, subject to error.

A study of the effectiveness of cooperative area anti-air warfare as a function of the quality of remote-to-local track correlation can be found in Lucas et al. (1990). By effectively coordinating weapon assignments, a battle group can ensure that the maximum number of threats is engaged. In stressful scenarios, a failure to coordinate can result in some threats being unintentionally over-engaged, while others are not engaged at all. This is clearly a function of raid count and is sensitive to correlation performance. Figure 6 shows how the number of leakers (to point defense) increases as a function of tracks per threat.

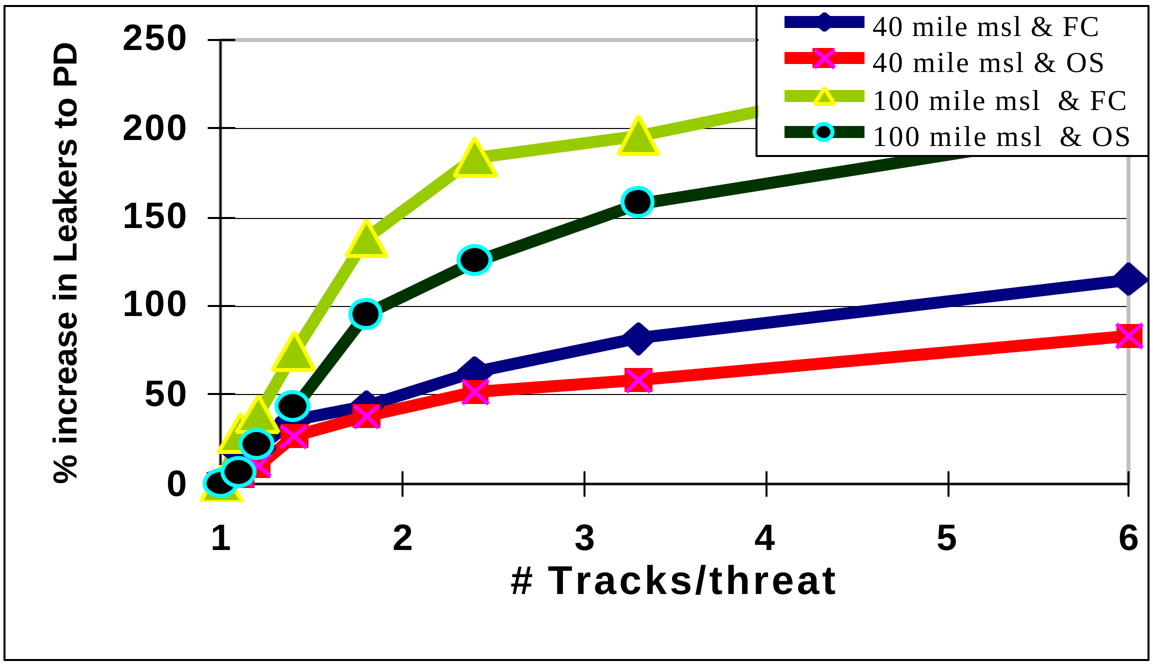

Figure 6: The sensitivity of leakers to the quality of data fusion for two missile ranges and weapon assignment algorithms (Lucas, Schwab, and Feldman [1990]).

The scenario consists of a defending Blue force with four area defense cruisers, one carrier, and one destroyer. These ships are attacked by a Red force firing 164 long-range cruise missiles from one threat axis within a span of five minutes. All six ships in the battle group exchange track data on a periodic communication link. The accuracy of the correlation decisions is varied from perfect correlation (one track per threat) to no correlation (up to six tracks per threat). It was believed *a priori* that the effect of data fusion would vary with weapon overlap and the degree of weapon assignment coordination. Consequently, two generic weapon ranges (40 and 100 miles) for the cruisers are investigated, as well as two different weapon assignment algorithms. One of the weapon assignment algorithms is force coordination (FC), and the other is ownship coordination (OS). In FC, a central algorithm makes all of the weapon assignments (other than emergency self-defense). In OS, each ship makes its own weapon assignments; however, they consider tracks' reported remote engagement status.

The number of leakers is sensitive to data fusion errors. The sensitivity increases significantly with weapon overlap and modestly to expanded coordination. Modern warfare places a premium on sharing battlefield information. Sharing inaccurate information, caused by data fusion errors, may well reduce effectiveness. Modeling the effectiveness of information systems without accounting for data fusion errors can badly bias results. The effectiveness of a deterministic algorithm that uses the average number of tracks per target will depend on the random distribution of the tracks per target. This was not examined in this study. If a simulation assumes perfect correlation, then, if the real correlation is not perfect, the estimated average number of leakers will be biased low.

Queues

Combat models contain many queues, such as communication and transportation networks, and items to be processed for acquisition and/or engagement. Because there is a great deal of literature on analyzing queues (e.g., Ross [1985]), we will address them only briefly. If items arrive at a server with a fixed time spacing and are serviced at a fixed service time, then the queue either remains empty (if the interarrival time is greater than the service time) or grows indefinitely. If the interarrival time is greater than the service time and only one of the components (arrival or service) is modeled stochastically, then the average queue congestion is biased (Hillestad and Moore [1996]).

For some special cases (e.g., independent homogeneous exponential arrivals and service times), there are closed-form solutions for the probability distributions and vital statistics (i.e., mean and variance) of queue congestion. These can be used to model stochastic effects with deterministic methods (see Hillestad [1982] and Hillestad and Carrillo [1980]). Unfortunately, in combat, events are quite complicated, with unpredictable batch arrivals, nonhomogeneous arrival rates, loss of server nodes, and time varying service priorities. Consequently, it is usually difficult to develop closed-form formulas that sufficiently characterize combat models' queues. Monte Carlo methods, while computationally intensive, can be used in these situations.

FINAL THOUGHTS

The inherent probabilistic nature of combat is…an issue of importance since the distribution of (potential) warfare outcomes and the divergence of time-phased processes are best studied [with methods that address stochastic issues]. Results based on deterministic models with expected values substituted for uncertain outcomes at each stage can be misleading—Bracken, Kress, and Rosenthal (1995, page 2)

Life is uncertain. Combat is too. Models that fail to properly account for combat's inherent randomness may produce misleading outputs. It is difficult, though not impossible, to quantify the effects of the stochastic nature of combat with deterministic models. Indeed, this is the preferred approach. However, the difficulty increases for large nonlinear simulations with state dependent decision logic. For the reasons articulated above, deterministic simulations should be used with caution. In particular, throughout any combat model, analysts must consider which events should be modeled stochastically. Attrition, detection, and battlespace awareness events should be given special scrutiny. When is this important and when is it not? Presently, we do not know the whole answer. However, the burden should be on those using deterministic approximations to justify their use. Random elements that are modeled via deterministic approximations should be tested for bias. Bayesian techniques are useful in quantifying uncertainty probabilistically, and Monte Carlo methods are effective in sampling from the vast ensemble of plausible model battles.

ACKNOWLEDGEMENTS

This paper is the product of over a decade of discussion on this topic with numerous individuals who have contributed to its content. Mark Youngren, Bart Bennett, and Rick Rosenthal have made particularly important contributions. In addition, referees' comments have improved the clarity and scope of this work.

REFERENCES

Allen, P., J. Gillogly, and J. Dewar, "Non-monotonic Effects in Models with Stochastic Thresholds," *PHALANX*, December 1993.

Ancker, C., "A Theory of Combat," in *Warfare Modeling*, Bracken, Kress, and Rosenthal (Eds.), Wiley, pp.165-198, 1995.

Ancker, C., and A. Gafarian, "The Validity of Assumptions Underlying Current Uses of Lanchester Attrition Rates," *Naval Research Logistics Quarterly*, Vol. 34, No. 3, 1987.

Bankes, S., *Modeling and Simulation Infrastructure to Support Adaptive Planning in the Air Force*, RAND, AB-125-AF, 1996.

Bennett, B., J. Hagen, J. Chow, A. Brooks, P. Myrick, W. Hobbs, and R. Gates, *Alternatives for Enhancing Aircraft Survivability: Phase 1*, RAND, MR-1003, 1998.

Berger, J., *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, 1985.

Bernardo, J., and A. Smith, *Bayesian Theory*, Wiley, 1994.

Berry, D., *Statistics: A Bayesian Perspective*, Duxbury Press, 1996.

Blackman, S., *Multiple-Target Tracking With Radar Application*, Artech House, 1986.

Box, G., "Robustness in the Strategy of Scientific Model Building," in *Robustness in Statistics*, Editors Launer and Wilkinson, Academic, page 202, 1979.

Bracken, J., M. Kress, and R. Rosenthal, "Introduction," *Warfare Modeling*, Wiley, 1995.

Clausewitz, C., *On War*, edited and translated by M. Howard and P. Paret, Princeton University Press, 1976.

Concepts Analysis Agency, CAA-TP-96-1, *Stochastic Concepts Evaluation Model (STOCEM)*, USA Concepts Analysis Agency, Bethesda, Maryland, August 1991.

Cleveland, W., *The Elements of Graphing Data*, Revised Edition, Hobart Press, 1994.

Cooper, G., "Non-Monotonicity and Other Combat Modeling Ailments," *PHALANX*, June 1994.

Davis, P., D. Gompert, R. Hillestad, and S. Johnson, *Transforming the Force: Suggestions for DoD Strategy*, RAND IP-179, 1998.

Davis, P., "Dynamic Instability," *PHALANX*, December 1992.

Defense Modeling and Simulation Office (DMSO), *Glossary of Modeling and Simulation (M&S) terms*, http://www.dmso.mil/, 1998.

Defense Science Board, *Report of the Task Force on Improving Test and Evaluation Effectiveness*, December 1989.

Dewar, J., S. Bankes, J. Hodges, T. Lucas, and P. Vye, *Credible Uses of the Distributed Interactive Simulation System*, RAND, MR-607-A, 1996a.

Dewar, J., J. Gillogly, and M. Juncosa, "Non-Monotonicity, Chaos, and Combat Models," *Journal of the Military Operations Research*, Vol. 2, Number 2, 1996b.

Dupuy, T., *Attrition: Forecasting Battle Casualties and Equipment Losses in Modern War*, Hero Books, 1990.

Dupuy, T., *Numbers, Predictions, & War*, Hero Books, 1985.

Efron, B., and R. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

Feller, W., *An Introduction to Probability Theory and Its Applications, Volume II*, 2nd Edition, Wiley, 1971.

Fox, D., and C. Jones, *JICM 3.0: Documentation and Tutorial*, RAND, DRU-1824-OSD, 1998.

Garner, J., "The Next Generation of Threat to U.S. Military Superiority…Asymmetric Niche Warfare," *PHALANX*, March 1997.

Good, I., "Some History of the Hierarchical Bayesian Methodology," in *Bayesian Statistics*, J. Bernardo et al., Editors, University Press, 1980.

Hammersley, J., and D. Handscomb, *Monte Carlo Methods*, Chapman & Hall, 1964.

Hillestad, R., and L. Moore, *The Theater-Level Campaign Model*, RAND, MR-388-AF/A, 1996.

Hillestad, R., J. Owen, and D. Blumenthal, "Experiments in Variable Resolution Combat Modeling," in *Warfare Modeling*, Bracken, Kress, and Rosenthal (Eds.), Wiley, pp. 63-86, 1995.

Hillestad, R., *Dyna-METRIC: Dynamic Multi-Echelon Technique for Recoverable Item Control*, RAND, R-2785-AF, 1982.

Hillestad, R., and M. Carrillo, *Models and Techniques for Recoverable Item Stockage When Demand and the Repair Process are Nonstationary*, RAND, N-1482-AF, 1980.

Hodges, J., and J. Dewar, *Is It You Or Your Model Talking? A Framework for Model Validation*, RAND, R-4114-A/AF/OSD, 1992.

Hodges, J., and R. Pyles, *Onward Through the Fog: Uncertainty and Management Adaptation in Systems Analysis and Design*, RAND, R-3760-AF/A/OSD, 1990.

Hoeber, F., *Military Applications of Modeling*, Military Operations Research, 1981.

Hofmann, H., R. Schnurer, and A. Tolk, "On the Impact of Stochastic Modeling in a Rule Oriented Combat Simulation Model on Division/Corps Level," in *Symposium on Coping With Uncertainty in Defence Decision Making, Technical Proceedings, Vol. II*, Editors Scheur, Faddy, and Wijnmalen, pp. 17.1-17.12, 1995.

Horrigan, J., *Aggregation, Chance, and Validity in Combat Simulation*, Horrigan Analytics, Technical Report, Chicago, IL, July 1967.

Huber, R., and A. Tolk, "Non-Monotonicity Effects in Combat Models and Tactical Decision Modeling," *PHALANX*, September 1994.

Hughes, W., *Military Modeling for Decision Making*, 3rd Edition, Military Operations Research Society, 1997.

Hughes, W., "A Salvo Model of Warships in Missile Combat Used to Evaluate Their Staying Power," in *Warfare Modeling*, Bracken, Kress, and Rosenthal (Eds.), Wiley, pp. 121-143, 1995.

Jaynes, E., *Papers on Probability, Statistics, and Statistical Physics*, R. Rosenkrantz, ed. Dordrecht, Holland: Reidel, 1983.

Johnson, R., E. Isensee, and W. Allison, "A Stochastic Version of the Concepts Evaluation Model (CEM)," in *Warfare Modeling*, Bracken, Kress, and Rosenthal (Eds.), Wiley, pp. 87-100, 1995.

Joint Chiefs of Staff, *TACWAR integrated environment*, *Executive Overview: Model Version 5.1*, TACWAR Configuration Control Group, TRAC, Fort Leavenworth, KS, 1998.

JWARS Office, Joint Warfare System (JWARS) Operational Requirements Document (ORD), http://www.dtic.mil/jwars/pdflib/ord_27.pdf, August 27, 1998a.

JWARS Office, *Joint Warfare System (JWARS) Overview briefing*, http://www.dtic.mil/jwars/pdflib/overview.pdf, August 26, 1998b.

Koopman, B., *Search and Screening*, Pergamon Press, 1980.

Koopman, B., "Fallacies in Operations Research," *Operations Research*, Vol. 4, Number 4, August 1956.

Lanchester, F., *Aircraft in Warfare: The Dawn of the Fourth Arm*, Constable and Co., Ltd., London, 1916.

Law, A., and W. Kelton, *Simulation Modeling & Analysis*, 2nd Edition, McGraw-Hill, 1991.

Louer, P., "More on Non-linear Effects...," *PHALANX*, March 1993.

Lucas, T., "How One Randomizes Matters: A Study of Non-monotonicity and Randomness in Combat Analysis," *PHALANX*, March 1997.

Lucas, T., K. Schwab, and S. Feldman, "The Impact of Data Fusion on Mission Effectiveness," *Technical Proceedings of the 1990 Joint Service Data Fusion Symposium*, Published by the Naval Air Development Center, 1990.

McCue, B., "Lanchester and the Battle of Trafalgar," *PHALANX*, December 1999.

Neimeier, H., "Analytic Uncertainty Modeling Versus Discrete Event Simulation," *PHALANX*, March 1996.

O'Bryon, J., *The Difficult Task of Striking a Balance Between Testing and Modeling and Simulation (M&S): The Live Fire Test Case*, http://www.dote.osd.mil/lfte/JOB97.HTM, 1996.

Oreskes, N., K. Shrader-Frechette, and K. Belitz, "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences," *Science*, Vol. 263, pp. 641-646, February 1994.

Press, S., *Bayesian Statistics: Principles, Models, and Applications*, Wiley, 1989.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edition, Cambridge, 1992.

Quade, E., *Analysis for Public Decisions*, 3rd edition, revised edition by G. Carter, North-Holland, 1989.

Ross, S., *Introduction to Probability Models*, 3rd Edition, Academic Press, 1985.

Royden, H., *Real Analysis*, 2nd edition, Macmillian, 1968.

Saeger, K., *Instabilities in Complex Deterministic Combat Models*, 68th MORS Symposium, June 2000.

Sandmeyer, R., *Simtech Project: Application of Supercomputers to Division/Corps Level Combat Simulation*, AMSAA, Interim Note No. C-159, October 1990.

Savage, S., *INSIGHT.XLA Business Analysis Software for Microsoft Excel*, Duxbury, 1998.

Science Applications International Corporation, *Integrated Theater Engagement Model (ITEM) Technical Manual*, Version 8.0, August 1998.

Stevens, W., "When Scientific Predictions Are So Good They're Bad," *New York Times*, September 29, 1998.

Taylor, J., *Lanchester Models of Warfare, Volume I*, Military Applications Section, Operations Research Society of America, 1983.

Thomas, C., "Models and Wartime Combat Operations Research," *Military Modeling for Decision Making*, 3rd Edition, edited by Hughes, Military Operations Research Society, 1997.

Tufte, E., *The Visual Display of Quantitative Information*, Graphics Press, 1983.

Welch, L., *Modeling and Simulation for Innovative Concepts*, Naval Postgraduate School Seminar, February 8, 1999.

Williams, J., *The Complete Strategyst*, Dover, 1986.

Wilson, B., and J. Hagen, *START Model Overview*, RAND, DRR-1529-1-AF, December 1996.

Yang, J., and A. Gafarian, "A Fast Approximation of Homogeneous Stochastic Combat," in *Warfare Modeling*, Bracken, Kress, and Rosenthal (Eds.), Wiley, pp. 357-387, 1995.

Youngren, M., *Uncertainty in Model Support to Decisionmaking*, personal communication of a working draft, Naval Postgraduate School, Monterey, CA, 1998.